(* TWO-WAY *)

22

Restrained Fits

(when numbers fade away)

		Exh.	Page
Introduction	22-1		
		A	
22A. Preparing an example	22-3	1	22-6/7
- singly restrained fit to all 36	22-5	2	22-8
		3	22-9/10
22B. A few holes need cause no trouble	22-15	4	22-11
		5	22-13/14
22C. More holes can matter	22-16	B	
		C	
22D. Doubly-restrained fits	22-23	6	22-18/19
- an inescapable example	22-33	D	
		7	22-24
22E. We can do still better	22-35	8	22-25
		9	22-26
22F. From the other extreme	22-44	10	22-28
		11	22-29
22G. Formulas can now help us (background)	22-54	12	22-30
		13	22-32
**********		14	22-34
		E	
22H. How far have we come?	22-61	15	22-36
		16	22-38/40
**********		17	22-42/43
		F	
22P. Problems for chapter 22	22-63	18	22-45
		19	22-46/47
		20	22-49
		21	22-50/52
		22	22-53
		G	
		23	22-55/57
		H	
		P	

some mathematical model far outside the area in which it was developed and tested. Whatever the character of the theory, we ought to be prepared for it to be right -- and equally for it to be wrong. Where "theory" suggests a restrained fit, we ought -- when we reasonably can -- to make at least two fits -- one restrained and one unrestrained -- and compare their quality.

Careful evaluation of the difference in quality of the two fits deserves concepts and sophistication that belong to confirmatory -- rather than exploratory -- data analysis. There will, of course, be situations where even the crudest look at the analyses tells us that the data either clearly seems to behave as predicted or utterly deviates from such prediction. In these cases, exploratory data analysis is enough. In others there will be an urgent need for confirmatory data analysis. In any event, it is good to begin by looking at the situation, whether or not the result of our look is crystal clear.

Restrained fits can also come from dire need, usually because the data is sparse. If only a few entries in a two-way table are present, we may not know enough to make a sensible two-way analysis. This is a common occurrence.

What sort of restrained fits are we likely to make? Once the versions of a factor, the things for which the fit to that factor is defined, come with decent numerical values, we can ask that our fit to them -- whether they label rows or they label columns -- be a straight-line fit. As always,

data appeared -- for the first time -- five years later in Johnson's 1949 textbook.

In this experiment, Johnson and Tsao Fei measured the sensitivity of various individuals to changes in pull on a ring hooked over one finger.

We shall be concerned with three of the many factors considered by Johnson and Tsao:

- the individual tested. Here we will look at only the four blind individuals called here **J, K, L,** and **M** --- IB(1), IB(2), IIB(1), and IIB(2) in Johnson's notation.

- three equally-spaced initial pulls --- weights --- called here **1, 4, 7,** as in Johnson's 1949 book and amounting to 100, 250, and 400 grams

- four equally-spaced rates of increase of pull --- rates --- called here **a, b, c, d,** as in Johnson's 1949 book and amounting to 100, 200, 300, and 400 grams per minute.

Both Johnson and Tsao Fei expressed the data in grams of increase of pull before the observer noticed the change. Let us start there. In fact, each situation was repeated five times for each individual on each of two days a week apart. Only the means of the sets of five similar trials are available.

For simplicity, then, we will begin with numbers found as follows:

mean of five similar trials on a day

sum of two such means, one for each day

exhibit 1 of chapter 22 -- finger sensitivity

Elementary analysis of 36 values
from the Johnson and Tsao Fei experiment
(entries are sums of two means)

A) The DATA

Weight & Subject	Rate			
	a	b	c	d
1K	39	85	101	151
4K	31	55	84	126
7K	26	56	70	98
1L	16	32	63	63
4L	12	22	33	51
7L	12	20	30	37
1M	18	31	42	58
4M	13	26	38	55
7M	14	26	40	46

B) The ELEMENTARY ANALYSIS -- by means

	a	b	c	d	eff	fit
1K	-28	-1	0	29	47	94
4K	-16	-11	4	22	27	74
7K	-9	-2	2	2	15	62
1L	5	2	-4	-3	-9	38
4L	9	0	-3	-7	-17	31
7L	14	3	-1	-16	-22	25
1M	8	2	-1	-7	-10	37
4M	7	1	-1	-6	-14	33
7M	9	2	2	-14	-15	32
eff	-27	-8	6	28	47	00
fit	20	39	53	75	00	-47

(Cont'd)

exhibit 2 of chapter 22 -- finger sensitivity

Diagnostic lot for lementary nalysis of exhibit 1

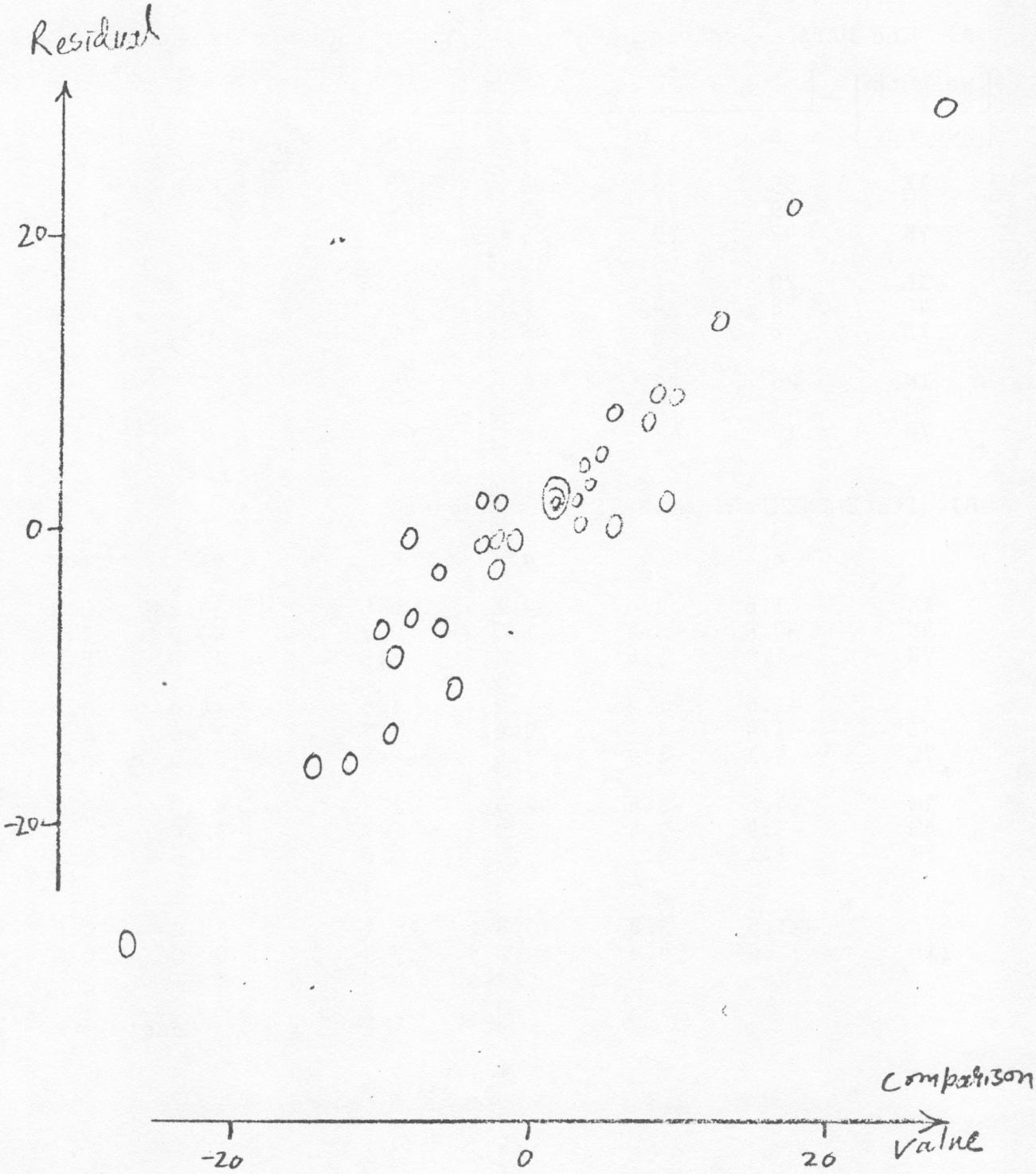


exhibit 3 of chapter 22 (cont'd.)

C) The RESIDUALS as a BATCH -- unit 0.1

```
to
 1    -4|8            16 + 3 + 17 = 36, hinges at depth 9h
 6    -3|44428        Hinges:          -17h and +15
 8    -2|03           Midspread:            32h
16    -1|64649023     Sides:           -50  and  47h
(3)   -0|444          Outside:               0
17     0|105         (combined par):        (2)
14     1|140505
 8     2|77
 6     3|65
 4     4|1647
```

D) The DATA as a BATCH -- unit 1.0

```
to
 1    11|8            17 + 5 + 14 = 36, hinges at depth 9h
 3    10|09           Hinges:          42  and  75h
 6     9|339          Midspread:            33h
 8     8|50           Sides:            8h and 109
13     7|45164        Outside:           2 and   1
17     6|3206        (combined par):       (2)
(5)    5|91287
14     4|929228
 8     3|40
 6     2|06
 4     1|15
 2     0|88
```

In panel A we have fitted a reasonable straight line to the "eff" and "fit" lines, using the simple procedure of chapter 9, which here amounts to putting a line through the two end values. The results are labeled "SL-fit" -- where "SL-" stands for "straight-line". The "SL-eff" values are found by subtracting 54.9, the mean of the SL-fits. The "Left over" line shows by how much the restrained effects fail to fit the free effects.

One sure and easy way to calculate the rest of the singly-restrained analysis is to use the "SL-eff" values to adjust the original data. This is done in panel B, and the resulting singly-restrained elementary analysis is set out in panel C.

The apparent quality of the fits is set out in panel E. Whichever of the last two columns we look at, we cannot avoid concluding that:

- the singly-restrained elementary analysis is much better than fitting only a mean
- the free elementary analysis is still a lot better than the singly-restrained elementary analysis.

As a result, we will want to make the free elementary analysis as we can. If we cannot, we will greatly prefer a singly-restrained elementary analysis to anything less detailed that we have yet seen.

Why might we not be able to use the free elementary analysis? Two reasons arise most often:

exhibit 5 of chapter 22 (Cont'd)

B) ADJUSTED VALUES -- data MINUS SL-eff

	a	b	c	d	mean
1K	87.5	102.5	90.5	89.5	92.5
4K	77.5	83.5	83.5	80.5	81.2
7K	70.5	84.4	75.5	70.5	75.2
1L	48.5	60.5	53.5	51.5	53.5
4L	36.5	43.5	42.5	42.5	41.2
7L	36.5	39.5	38.5	28.5	35.8
1M	54.5	58.5	52.5	47.5	53.2
4M	39.5	51.5	48.5	45.5	46.2
7M	43.5	51.5	50.5	37.5	45.8
				(mean)	(58.3)

Examples = 87.5 = 59 - (-28.5), 102.5 = 93 - (-9.5),
90.5 = 100 - (9.5), 89.5 = 118 - (28.5).

C) SINGLY-RESTRAINED ANALYSIS -- restrained residuals

	a	b	c	d	eff	fit
1K	-5.0	10.0	-2.0	-3.0	34.2	92.5
4K	-3.7	2.3	2.3	-0.7	22.9	81.2
7K	-4.7	9.2	0.3	-4.7	16.9	75.2
1L	-5.0	7.0	0.0	-2.0	-4.8	53.5
4L	-4.7	2.3	1.3	1.3	-17.1	41.2
7L	0.7	3.7	2.8	-7.3	-22.5	35.8
1M	1.3	5.3	-0.7	-5.7	-5.1	53.2
4M	-6.1	5.3	2.3	0.7	-12.1	46.2
7M	-2.3	5.7	4.7	-8.3	-12.5	45.8
SL-eff	-28.5	-9.5	9.5	28.5	58.3	00.0
SL-fit	29.8	48.8	67.8	86.8	00.0	58.3

Examples: -5.0 = 87.5 - 92.5, 10.0 = 102.5 - 92.5, etc.

(Cont'd)

- the results of the analysis are wanted as part of some larger analysis, where only very simple results can be conveniently used. The larger analysis may be a matter of arithmetic, when the need for restraint is usually clear. It may also be a matter of how the ultimate recipients think about some more complex problem, only part of which is illuminated by the body of data being analyzed. Here the need for restraint is usually less clear -- indeed, it is often a matter of debate.

- there may be many absent values in the two-wau table.

We turn next toward examples where the second reason applies. Parts of the example just considered will serve.

22B. A few holes need cause no trouble

We have already become used, especially in chapter 20, to making two=way analyses of tables with quite a few holes. Any reader should be able to drop out 5 or 10 of the 36 values in exhibits 1 or 3 and then carry out an unrestricted two-way analysis on the remainder. To cause us to change to a restricted analysis requires a more violent dose of absentness, either many individual absences or an uncomfortable pattern of absences.

Symmetry suggests that the natural place to take the intercept is halfway between b and c, so that it corresponds to what would be the row adjustment for the full table. To keep mental arithmetic simple, we can think of the following assignment of working values:

a	b	c	d
-1.5	-0.5	0.5	1.5

whose unit spacing of course leads to

tilt = 1/2(value for c) - 1/2(value for a) column (**)

and in

tilt = 1/2(value for d) - 1/2(value for b) exhibit 6

Only a little thought is needed to see that

intercept = 3/4(value for c) + 1/4(value for a)

and (*)

intercept = 3/4(value for b) + 1/4(value for d)

are the formulas we need for the values in column (*) of exhibit 6.

The differences between these intercepts and the row fits of exhibit 5 are not very large. In stem-and-leaf form they are, in units of 0.1:

```
 4|8
 3|81
 2|
 1|61
 0|
-0|22
-1|0
-2|6
```

Hinges: -0.2 and +3.1
Midspread: 3.3
Range: 56.7 or 53.0

exhibit 6 (cont'd.

B) ELEMENTARY ANALYSES of 5x2 and 4x2 subtables -- by means

	a	c	eff	fit
1K	0.4	-0.4	28.6	79.5
7K	-0.6	0.6	12.6	63.5
4L	-1.1	1.1	-20.9	30.0
1M	2.9	-2.9	-6.9	44.0
7M	-1.6	1.6	-13.4	37.5
eff	-20.9	20.9	50.9	00.0
fit	30.0	71.8	00.0	-50.9

	b	d	eff	fit
4K	-2.1	+2.1	26.9	91.5
1L	+0.9	-0.9	0.9	65.5
7L	+1.9	-1.9	-21.1	45.5
4M	-0.6	+0.6	-6.6	58.0
Eff	-15.4	15.4	64.6	00.0
fit	49.2	80.0	00.0	-64.6

alone -- or the one that corresponds to column adjustments alone -- to take any shape the data may seem to call for. If nine row summaries run -- as in exhibit 3 -- 92,81,75, 53,41,36,53,46,46, we accept this sequence almost as happily as any other. So far as among-rows or among-columns pieces are concerned, such an elementary analysis of a two-way table is thoroughly free-form. (Its fixed-form characteristics are confined to its insistence on the use of -- and combination of -- row effects and column effects in a particular and essential way.)

The way we were able to deal with the quite scattered data of exhibit 6 was exactly to replace this free-form freedom by fixed-form rigidity. We decided that we would have to treat the effect of rate as a straight-line dependence. This additional amount of rigidity was quite enough to let us do a meaningfully effective analysis.

We ought to take this as a sign and a portent. As we deal with more and more scattered data we must expect to have to use more and more rigidly fixed forms of partial description. This must happen -- there is no escape.

Before we close this section, there is one question we should not fail to answer: What corresponds to residuals in the analysis we have made? One simple reasonable answer is: The differences between individual tilts and the mean of these tilts. These differences are, in order: +2.1, -0.9, +3.1, -3.9, +3.6, -4.9, -0.4, -2.4, +4.1, as we can see from panel A of exhibit 6.

of choosing a shape for this behavior, we will gain better intercepts, and, probably, better residuals.

22D. Doubly-restrained fits

As we faced more and more absent values, we found it desirable -- even necessary -- to restrain our fit to one of the factors. What if even more values are absent?

Clearly we may want to restrain our fit to both factors.

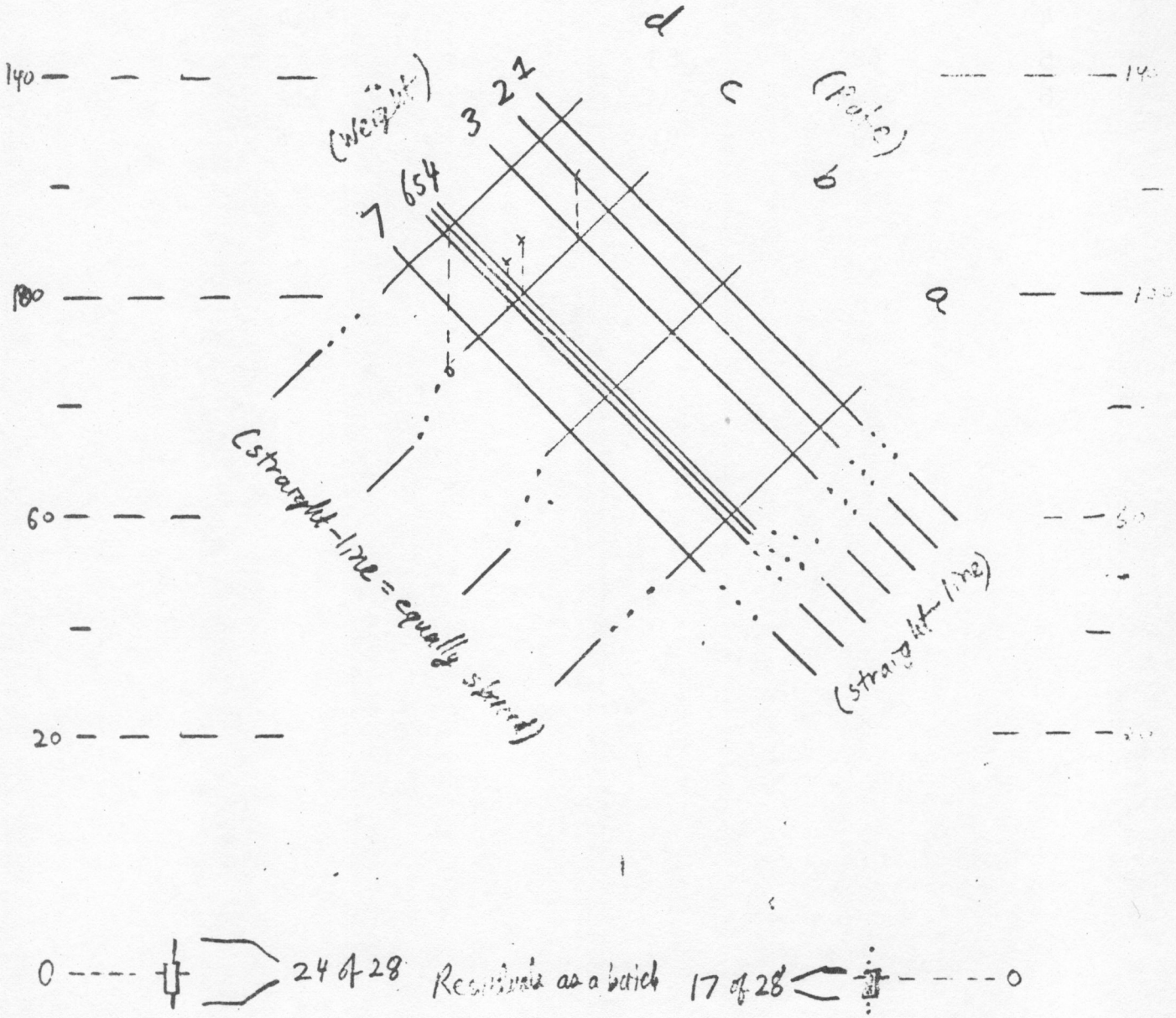
Of course we can only do this if both factors have naturally numbered versions. The example we have been using thus far in this chapter has naturally numbered versions for the rate factor. The versions of the other factor -- combining subject and weight -- do not seem to be numbered in any reasonably natural way. Let us, then, choose a different example.

Exhibit 7 presents a different slice out of the Johnson and Tsao Fei data, involving a fourth blind subject. The elementary analysis, carried out in detail in the previous chapter, shows quite small residuals, with a few exceptions. Exhibit 8 two-way plots the fit, showing that the dependences on both rate and weight are fairly nearly straight-line. The most noticeable deviations from equal-spacing are a tendency for curvature against rate and a displacement of the fit for weight 4 close to the fit for weight 5.

Exhibits 9 and 10 give selected portions of exhibit 7, chosen to raise a variety of questions. Exhibit 9 offers a close analog to exhibit 6. Again it would seem unwise

exhibit 8 of chapter 22 -- finger sensitivity.

The fit of exhibit 7 and the straight-line fits



to fit straight-line behavior across rate. This time, however we can still get by with half-measures, because we can rather reasonably fit straight-line behavior in weight. (Details are left to the reader.)

The same is almost true of the seven data points that survive in exhibit 10. We can fit straight-line behavior in weight for rates a and d easily, for rate b somewhat less securely, and for rate c not at all. However, if we use the intercept at weight 4 as one constant in our straight-line fit, we can find that constant for rate c. The fact that we cannot find -- and do not need to use -- a tilt against weight for rate c need not concern us.

The fit to exhibit 10 thus produced takes the form shown in exhibit 11. Since there were never more than two data points for any rate, a separate straight line for each rate -- with its two constants -- fits the data exactly, leaving all residuals zero. Instead of a partial description, we have obtained a total one.

If we wish to obtain a partial description, we must restrict the form of our fit somewhat further. The most reasonable simple single restriction proves to be a requirement that all the lines relating response to weight shall have the same slope.

The results of such a fit are shown in exhibit 12. The slopes in exhibit 11, as measured by the difference in fitted response between weight 1 and weight 7, were 33, 18, a, and 27. Of these the first is the firmest, at least so

exhibit 11 of chapter 22 -- finger sensitivity

The fit to exhibit 7 by straight-line behavior against weight, separately for each rate

A) The FIT

Weight	Rate			
	a	b	c	d
1	82.	97	a	134.5
2	76.5	94	a	130
3	71.	91	a	125.5
4	65.5	88	110	121
5	60	85	a	116.5
6	54.5	82	a	112
7	49	79	a	107.5

B) The RESIDUALS

Weight	Rate			
	a	b	c	d
1	0.0	a	a	a
2	a	a	a	0.0
3	a	0.0	a	a
4	a	a	0.0	a
5	a	0.0	a	a
6	a	a	a	0.0
7	0.0	a	a	a

far as the spacing of its observations go. Carrying this thought further leads to taking

$$\frac{6(33) + 2(18) + 0(x) + 4(27)}{6 + 2 + 0 + 4} = 28.5$$

as a reasonable summary. Since this is for the six steps from 1 to 7, we fit a downward slope of 28.5/6 = 4.8 per step.

We now have a rather reasonable fit (partial description), which we can describe by giving:

- the common slope of -4.8 (per step in weight).
- the intercepts (at weight 4) of 65.5, 88, 110, and 121 (for rates a, b, c, and d, respectively).

If we want to compress this (now) partial description somewhat further, we can try to replace the four intercepts by some more compact description. While the behavior of the intercepts is not quite that of points along a straight line, straight-line behavior describes the four intercepts quite well. A simple approach gives 67.4, 86.2, 105.0, 123.8 as a straight-line fit, so that the situation is now in the form shown in exhibit 13. Clearly a uniform decrease of one to two units in the fitted values will be helpful, making the residuals somewhat smaller as a whole.

We have now a quite condensed and convenient partial description of the seven values present in exhibit 14, a description:

- involving three constants (one general level and two slopes).
- thus of a very fixed form.

As the amount of data -- and the degree of balance -- becomes less and less, we have found it natural to move more and more toward a fixed-form fit. Indeed, at this last step, we have adopted a wholly fixed form. As we said above, this is inevitable. Though inevitable, it still deserves emphasis.

* an inescapable example *

To some, our examples of more and more sparse two-way tables may seem quite inrealistic. As an antidote, let us turn back to chapter 14, where we fitted both tilt and bend, in particular with

$$y = \text{value of U.S. imports}$$

and

$$x = \text{date}$$

There we thought of y plotted against x and fitted a quadratic curve in the xy-plane. That approach met our needs as we then saw them. It is time to take another.

It will be convenient here to take the fit in the form

$$a + bx + c\left((x-1923.5)^2 - 0.25\right)$$

treating $(x-1923.5)^2 - 0.25$, for all that it is determined once x is determined, as if it were the numerical value, t, describing a separate factor.

The 12 years of data on U. S. imports now appears as in exhibit 14. As a two-way table this mainly consists of absent values, as is both inevitable and striking. Looking at this table, we get a different view of bend and tilt finding than we had in chapter 14.

So far we have chosen 1, x and $(x-1923.5)^2$ -- the subtraction of 0.25 makes no real difference -- as the carriers of our straight-line dependence -- as the quantities that we are to multiply by freely choosable coefficients, adding the results together to give our fit, our incomplete description. With this choice we have a very sparse two-way table, but we expect no trouble in fitting. In exhibit 14, b x and c $((x-1923.5)^2 - 0.25)$ behave quite differently. As a consequence there is little possibliity of confusing one for another -- and there is little arithmetical difficulty in fitting both -- either simultaneously or one after the other.

22E. <u>We can do still better</u>

We jumped into the analysis of the last few sections.

The careful reader will have noticed that although choice of expression for the numbers associated with the versions of a factor was discussed in the introduction, we have done nothing to make such a choice. We knew that rates **a**, **b**, **c**, and **d** corresponded to 100, 200, 300, and 400 grams a minute. We took this as reason to take these four rates as equally spaced without looking to see whether we could do better. This was rather uncalled for.

Were we to plot fitted effect or fitted adjustment (= fitted central value) against rate, we would find a curve convex toward increasing delay but concave toward increasing rate. If we are to straighten out this plot by re-expressing rate, we have thus to move down the ladder of powers. Exhibit 15 explores the possibilities. Clearly logs should do well for us.

Exhibit 16 deals with the corresponding singly-restrained fit. We find that this fit is almost as good as the unrestrained two-way analysis.

More than this can be said. Turning back to panel A of exhibit 20, we see that the slope of the fitted straight-line effects is

$$\frac{24.2 - (-32.6)}{60.2 - 0.0} = \frac{56.8}{60.2} = .94$$

Thus we can put ~~7~~ with decreasing quality of approximation -- what we see about rate as follows:

Free analysis in logs:	-31.8, -3.8, 10.8, 25.1
SL-analysis in logs:	slope = 0.94
same, rounded:	slope = 1

To an approximation of intermediate quality only, then, we have that the response

100 log delay (measured in grams)

is fitted by

100 log rate + something depending on weight and subject.

To the thoughtful reader, this says much that the free fit hardly hinted at -- and the unrounded singly-restricted fit had not made crystal clear:

- if the delay is measured in grams, it is nearly proportional to the rate in grams per minute. Therefore the delay, for a given weight and subject, is nearly constant in minutes.

Clearly we ought, if we are seriously interested in understanding this data, to go back and start again, expressing the response as

exhibit 16 of chapter 22 (Cont'd)

C) SINGLY-RESTRAINED ANALYSIS

	a	b	c	d	eff	fit
1K	-0.9	4.7	-5.0	1.3	34.2	92.5
4K	0.3	-3.1	-0.8	3.5	23.0	81.3
7K	-0.7	3.9	-2.8	-0.5	17.0	75.3
1L	-0.8	1.8	-2.9	2.1	-4.9	53.4
4L	-0.6	3.0	-1.7	5.3	-17.1	41.2
7L	5.0	-1.4	-0.1	-3.1	-22.7	35.6
1M	-5.6	0.2	-3.5	-1.2	5.3	53.0
4M	-2.7	-0.1	-0.8	3.5	-12.0	46.3
7M	1.8	0.4	-8.3	-4.0	-12.5	45.8
SL-eff	-32.6	-4.2	12.5	24.2	58.3	00.0
SL-fit	25.7	54.1	70.8	82.5	00.0	58.3

D) SINGLY-RESTRAINED RESIDUALS as a BATCH -- unit = 0.1

to		
1	-8	3
	-7	
	-6	
2	-5	0
3	-4	0
6	-3	151
9	-2	789
12	-1	472
(9)	-0	978618185
15	0	324
12	1	883
9	2	1
8	3	9055
4	4	7
3	5	063

12 + 9 + 15 = 36, hinges at depth 9h

Hinges:	-22	and	+19h
Midspread:			41h
Sides:	-63h	and	+61
Outside:	1	and	0
(combined par):			(2)

log of delay in time rather than as log of delay in weight. When we do, things will be even simpler.

Turning next to exhibit 10 (section 22C) where we could easily fit one 5-by-2 and one 4-by-2 table, we find the tilts against log rate to be

$$\text{a to c} \qquad \frac{20.9 + 20.9}{.477 - .000} = 88$$

$$\text{b to d} \qquad \frac{15.4 + 15.4}{.602 - .301} = 101$$

whose ratio is much nearer unity than was the case for the tilts of 20.9 and 15.4 when rates were used. Thus even the 18 data values of exhibit 10 were enough to urge us to use log rate rather than raw rate in our restrained analysis.

The result of seeking out tilts and intercepts -- the latter now at

$$(1/4)(.000 + .301 + .477 + .602) = .345$$

is set out in exhibit 17. The overall tilt is between 94 and 91, to be compared with 94 from the whole 9x4 table in exhibit 10. The fitted intercepts for weight-subject combinations show the same general pattern as for the whole data in exhibit 10, only one value being shifted by more than 2.1 (and that one by only 4.0).

Using log rate again gives a quite satisfactory analysis.

exhibit 17 of chapter 22 (Cont'd)

B) TILTS and INTERCEPTS

	Intercept	Tilt	(*)
1K	88.5	86	30.9
4K	79.2	116	21.6
7K	75.8	90	18.2
1L	55.1	96	-2.5
4L	39.7	92	-17.9
7L	34.0	89	-23.6
1M	51.9	76	-5.7
4M	46.8	106	-10.8
7M	47.4	94	-10.2
	(mean) (57.6)	94	
	(median)	92	

Examples: 88.5 = .28 (59) + .72 (100) = 100 - .28 (100-59)
79.2 = .85 (74) + .15 (109) = 79 + .15 (109-74)
86 = 2.1 (100) - 2.1 (59) = 2.1 (100-59)
116 = 3.3 (109) - 3.3 (74) = 3.3 (109-74)

(*) = intercept MINUS mean intercept, compare with column "fits" of exhibit 10, and with column (*) of exhibit 8.

exhibit 18 of chapter 22 -- finger sensitivity

A fragment of exhibit 11 that is still more obviously a problem

Weight	Rate			
	a	b	c	d
1	a	a	a	134
2	a	a	a	a
3	a	a	122	a
4	a	a	a	a
5	a	85	a	a
6	a	a	a	a
7	49	a	a	a

exhibit 19 of chapter 22 (Cont'd)

C) Some things the FIT MIGHT BE -- for the full table

	Rate tilt				Weight tilt			
	a	b	c	d	a	b	c	d
1	53.7	82.9	112.1	141.3	141.3	141.3	141.3	141.3
2	53.7	82.9	112.1	141.3	126.7	126.7	126.7	126.7
3	53.7	82.9	112.1	141.3	112.1	112.1	112.1	112.1
4	53.7	82.9	112.1	141.3	97.5	97.5	97.5	97.5
5	53.7	82.9	112.1	141.3	82.9	82.9	82.9	82.9
6	53.7	82.9	112.1	141.3	68.3	68.3	68.3	68.3
7	53.7	82.9	112.1	141.3	53.7	53.7	53.7	53.7

29.2 for each horiz. step — 0.0 for each horiz. step

0.0 for each vert. step — 14.6 for each vert. step

	Mixed tilt				Another mixture			
	a	b	c	d	a	b	c	d
1	111.9	121.7	131.5	141.3	171.3	161.3	151.3	141.3
2	102.2	112.0	121.8	136.6	151.7	161.7	131.7	121.7
3	92.5	102.3	112.1	121.9	132.1	122.1	112.1	102.1
4	82.8	92.6	102.4	112.2	112.5	102.5	92.5	82.5
5	73.1	82.9	92.7	102.5	92.9	82.9	72.9	62.9
6	63.4	73.2	83.0	92.8	73.3	63.3	53.3	43.3
7	53.7	63.5	73.3	83.1	53.7	3.7	33.7	23.7

9.8 for each horiz. step — -10.0 for each horiz. step

9.7 for each vert. step — 19.6 for each vert. step

exhibit 20 of chapter 22 -- finger sensitivity

The example of exhibit 18 reconsidered

A) The DATA -- with rate expressed as 1, 3, 5, and 7

Weight	Rate			
	1	3	5	7
1	a	a	a	134
2	a	a	a	a
3	a	a	122	a
4	a	a	a	a
5	a	85	a	a
6	a	a	a	a
7	49	a	a	a

B) In terms of SUM and RATE

Sum	Rate			
	1	3	5	7
8	49	85	122	134

exhibit 21 of chapter 22 (Cont'd)

C) FIT and RESIDUALS -- weight tilt only

	Fit				Residuals			
	a	b	c	d	a	b	c	d
1	141.4	141.4	141.4	141.4	a	a	a	-7.4
2	126.2	126.2	126.2	126.2	a	a	a	a
3	111.0	111.0	111.0	111.0	a	a	11.0	a
4	95.8	95.8	95.8	95.8	a	a	a	a
5	80.6	80.6	80.6	80.6	a	4.4	a	a
6	65.4	65.4	65.4	65.4	-8.4	a	a	a
7	50.2	50.2	50.2	50.2	a	aa	a	a

0.0 for each horiz. step
15.2 for each vert. step

range = 19.4
midspread = 15.6

D) FIT and RESIDUALS -- a mixed tilt

	Fit				Residuals			
	a	b	c	d	a	b	c	d
1	-22.8	30.8	84.4	138.0	a	a	a	-4.0
2	-7.6	46.0	99.6	153.2	a	a	a	a
3	7.6	61.2	114.8	168.4	a	a	9.2	a
4	22.8	76.4	120.0	183.6	a	a	a	a
5	38.0	91.6	145.2	198.8	a	-6.6	a	a
6	53.2	106.8	160.4	214.0	3.8	a	a	a
7	68.4	122.0	175.6	229.2	a	a	a	a

53.6 for each horiz. step
-15.2 for each vert. step

range = 13.8
midspread = 11.1

(Cont'd)

exhibit 22 of chapter 22 -- finger sensitivity

A larger fragment of exhibit 11

Weight	rate			
	a	b	c	d
1	a	a	a	134
2	a	a	112	130
3	a	a	122	a
4	a	84	110	a
5	a	85	a	a
6	57	81	a	a
7	49	a	a	a

exhibit 23 of chapter 22 -- finger sensitivity

Fitting a larger fragment

A) DATA, with RATES expressed in SIMPLE NUMBERS

Weight	Rate			
	1	3	5	7
1	a	a	a	134
2	a	a	112	130
3	a	a	122	a
4	a	84	110	a
5	a	85	a	a
6	57	81	a	a
7	49	a	a	a

B) DATA in terms of SUM and RATE

Sum	Rate			
	1	3	5	7
7	57	84	112	a
8	49	85	122	134
9	a	81	110	130

Sum = weight + rate

(Cont'd)

exhibit 23 of chapter 22 (Cont'd)

E) STRAIGHT-LINE ANALYSIS for ORIGINAL CARRIERS

Weight	Rate				SL-eff	SL-fit
	1	3	5	7		
1	a	a	a	-2	-3	93
2	a	a	5	-7	-2	94
3	a	a	14	a	-1	95
4	a	5	1	a	0	96
5	a	5	a	a	1	97
6	6	0	a	a	2	98
7	-3	a	a	a	3	99
SL-eff	-45	-15	15	45	94	00
SL-fit	49	79	109	139	00	-94

Notes.

1) Residuals are the saemaas in panel D.

2) Weight SL-eff: +1 per step, same as SL-eff in panel D for sum.

3) Rate SL-eff: +30 per step = 1·(28 per step) + 2·(1 per step)
where

SL-eff in panel D = +28 per step for rate

and

SL-eff in panel D = +1 per step for sum.

Response is partly described by $a + f(\text{rate}) + g(\text{wt})$

or:

Response is fitted by $a + f(\text{rate}) + g(\text{wt})$

In the developments of sections 22C and 22D, we dealt with increasing absence of data by specializing the form of dependence of our fits on rate and weight. We could have jumped from

$$g(\text{wt})$$

to the much more restricted

$$c_1 \cdot (\text{"wt"}) + \text{a constant}$$

where c_1 is a constant to be fitted and "wt" takes the values 1,2,3,4,5,6, and 7.

Actually we made one intermediate stop, but by exhibit 16 we had completed the jump, reaching

Response is fitted by $A + f(\text{rate}) + c_1 \cdot (\text{"wt"}) + \text{constant}$

As more observations faded away, we could not maintain even this limited generality, and had to replace the general

$$f(\text{rate})$$

by the particular

$$b_1 \cdot (\text{"rate"}) + \text{another constant}$$

****

22H. How far have we come?

In this chapter we have been concerned with what we can do with two-way tables in which some or many values are absent. "What we can do" has here two meanings:

- What sort of fits is it wise to undertake?
- How can and should we set about making the fits?

The first of these questions has a clear and simple set of answers. If enough of the two-way table is present, we can go ahead with any sort of analysis by differences, particularly analysis by median differences.

If there are not enough entries to make such an analysis easy and effective, we have either to make use of more structure in the data, or to give up the idea of a fit that involves both factors.

The simplest way to utilize structure in the data arises when it is reasonable:

- to attach numerical values to the various version of one or both factors,

and

- to restrict the dependence of the fit on that factor to a straight-line dependence in terms of these numerical values. (As is so often the case, choice of expression -- both for factors and responses, can clearly be important.)

The more fragmentary the data, the more firmly we are forced to adopt more and more rigid restrictions on our fit.

As an example has shown us, data can be so fragmentary as to make the fitting of even a straight-line dependence on

****

22P. Problems for chapter 22

1. Do the problem of exhibit 19.

2. Do the problem of exhibit 21.